

Early Grade Reading Assessment (EGRA): Cheat Sheet

Orally administered

Assesses building blocks of reading and comprehension

Administered in languages children read and understand

What it is. EGRA measures reading comprehension and the precursor skills that lead up to it (listening comprehension, letter-sound knowledge, decoding, fluency). An adaptable framework to any language, context, and purpose, EGRA provides information on each skill that constitutes reading ability.

EGRA Subtasks measure precursor and comprehension skills to build a full picture of reading ability

EGRA is designed to assess each skill a child needs to read. EGRA starts with precursor skills and builds to reading comprehension, with each subtask providing new information about the reading skills a child does and does not (yet) have.

COMPREHENSION

Reading comprehension

Adds: drawing out meaning from connected text

Understand what they read

Sam was lost in the market. He opened his map and turned it around. Then he smiled and walked to the big tree.

DIRECT What did Sam open?
INFERENCE Why did Sam smile?

MEASURED WITH

sentence verification, passage reading comprehension, maze.

WITHIN THIS SKILL

Comprehension is measured after reading a sentence or passage and asking questions to assess meaning-making of what they read.

WHY IT MATTERS

Oral language, decoding, and fluency all contribute to it directly.

Fluency

Adds: accuracy and automaticity of reading connected text

Read text aloud accurately, at a sufficient rate

READS ALOUD

The man has a map. He uses the map to find his way to the main road in town. As he walks through the streets, he sees the landmarks shown in his map. He is surprised when...

MEASURED WITH

reading accuracy, oral reading fluency.

WITHIN THIS SKILL

Oral reading fluency is timed (typically one minute) and scored in correct words per minute. Giving readers enough time to read to the end allows assessing comprehension using the same passage.

WHY IT MATTERS

Words must be read at a pace that makes meaning-making possible.

Alphabetic principle

Adds: print, then whole words

PART 1 · LETTER-SOUND CORRESPONDENCE (PHONICS)

Match each written letter to its sound or name

m → /m/ its sound

PART 2 · DECODING

Use letter sounds to read whole words

m **a** **p** → map

MEASURED WITH

letter sound, letter name, and syllable identification; dictation.

WITHIN THIS SKILL

In many languages, letter sounds are more directly linked to decoding than letter names.

WHY IT MATTERS

Knowing the sound a letter represents is what lets a learner sound out a word.

MEASURED WITH

familiar word reading, nonword reading.

WITHIN THIS SKILL

In some contexts, familiar words may be recognized instantly without sounding out letters; while nonwords, like *vap*, must be sounded out.

WHY IT MATTERS

Reading whole words is how sound and letter knowledge feed into comprehension.

Phonological awareness

Adds: focus on recognizing individual sounds

Hear the separate sounds in spoken words

HEARS "map" → /m/ first sound

MEASURED WITH

initial sound identification, letter sound discrimination, phoneme (or syllable) segmentation.

WITHIN THIS SKILL

Recognizing specific sounds that make up a word either at the beginning or (more difficult) in the middle or end of a word.

WHY IT MATTERS

Children who start school with little of it have trouble acquiring alphabetic coding skills.

Oral language

Comprehend and understand discourse

Understand spoken language and know word meanings

HEARS "map" →

sun map cup glasses

Receptive vocabulary: picks the map from four pictures

MEASURED WITH

listening comprehension; new RELM picture matching, short story, receptive and expressive vocabulary, semantic fluency.

WITHIN THIS SKILL

Easiest task of the rest and should come first and assesses comprehension independent of reading skills.

WHY IT MATTERS

Understanding spoken language contributes directly to reading comprehension, and works together with the other precursor skills to support reading.

PRECURSOR SKILLS

Examples are illustrative, in English; EGRA subtasks are adapted to each language. RELM = Receptive and Expressive Language Module. Source: EGRA Toolkit 3.0.

01 DESIGN YOUR EGRA decide why, for whom, and how

Purpose	Grade / method	Key characteristics
Inform classroom instruction (formative)	Census or sample-based study. Must time it to the academic calendar and curriculum.	Results intended for teachers first, including support to understand and use the results to inform instruction. Frequency, who conducts assessment, drives cost.
Assess progress, report to national or global indicators (summative)	Typically measured using a nationally representative sample of learners, global SDG 4.1.1a indicator is measured at Grade 2 or 3 .	Consult UIS criteria and data review requirements for SDG 4.1.1a reporting, select subtasks to meet reporting and national tracking purposes, benchmark needed
Measure progress over time (evaluative)	Longitudinal tracking of cohorts of learners or schools, transition across curriculum or grades, treatment vs comparison schools, or census.	Typically relies on comparability based on equivalent test forms, may require larger sample, benchmark needed.

NOT FOR High-stakes decisions (e.g. grade promotion), evaluating an individual teacher, or as an instructional program in itself.

DESIGN REMINDERS

Universal Design for Assessment: inclusive population, precisely defined constructs, accessible non-biased items, amenable to accommodations, simple instructions, readable and legible. Build it in from the start.

Grade range: EGRA reliably measures up to Grade 4. Go higher for below-standard readers, crisis-interrupted learning, overage or out-of-school learners, second-language learners, or cohort tracking.

EGRA TEAM ROLES

Role	Does	Role	Does
EGRA Manager	Runs process and budget	UDA Experts	Inclusive assessment design
Reading Specialists	Item review; interprets results	Psychometricians	Item analysis; quality criteria
Language Experts	Translation; multi-language use	Statisticians / Analysts	Sampling, weights, analysis
Curriculum Experts	Aligns items to curriculum	Sign / Braille Experts	Adapts for deaf and blind learners

EGRA SUBTASK MENU what to measure, how many items, when to stop · aim for 5–8 subtasks, 15–25 min total

Domain	Subtask	Type	Min. # of items	Early stop	Pause rule	Notes
Oral language	Listening comprehension	COMMON	10	None	10 sec	Untimed; 2 short passages. Often first or last.
Phonological awareness	Initial sound identification	COMMON	20	5	3 sec	Untimed; one-syllable words.
	Letter sound discrimination	COMMON	10	n/a	3–5 sec	Untimed; practice sets first.
	Phoneme segmentation	ADDITIONAL	10	5	3 sec	Floor effects; use if ISI shows ceiling.
Alphabetic principle (letter-sound)	Letter sound identification	COMMON	50	10	3 sec	1 min; letter-sound correspondence.
	Letter name identification	COMMON	50	10	3 sec	1 min. Ceiling effects common.
	Syllable identification	COMMON	50	10	3 sec	1 min. Open-syllable languages.
	Dictation	ADDITIONAL	10	None	15 sec	Hard to score in a standard way.
Alphabetic principle (decoding)	Familiar word reading	COMMON	50	5	3 sec	1 min. Frequent grade-level words.
	Nonword reading	COMMON	50	5	3 sec	1 min. Tests pure decoding.
Fluency	Oral reading	COMMON	2–3 passages	None*	3 sec	Timed at 1 minute and allow to read to end if linked to RC; fluency = correct words/min.
Reading comprehension	Sentence verification	COMMON	5–10	None	10 sec	True/false per sentence.
	Passage reading comprehension	COMMON	10+ (based on 2+ passages)	None	10 sec	Mix of direct and inferential questions of varying difficulty; no more than 5 RC questions per reading passage.
	Maze	ADDITIONAL	10	n/a	n/a	Pick the word that fits from three.

An early stop scores zero on that subtask and moves on; an assessment auto-stop can end the whole EGRA after early stops on the first 2–3 reading subtasks. After a pause, mark the item incorrect and say “go on”. Passage comprehension: 4–5 items per passage. ISI = initial sound identification. *For oral reading, early stop rule at 5 consecutive wrong words may be used when the passage is not also used for comprehension.

02 ADAPT fit the EGRA framework to your context

- Remember to always adapt and contextualize subtasks to your purpose and population of learners.
- When no previous tools exist, consider first completing a language analysis, curriculum review. Always field test items for sufficient quality and differentiating ability with a comparable group of learners.
- Plan early for an adaptation workshop: about 5–6 days and 30 participants (officials, subject experts, teachers, a psychometrician), with at least two facilitators. Start story writing early and include pretesting.
- Check face and content validity in the workshop; construct validity comes from pilot data. Translate learner instructions rather than rewriting them.

EGRA DEVELOPMENT: ACCOMMODATION VS MODIFICATION

An **accommodation** changes how a learner is tested, not what is tested (larger font, sign-language instructions, quiet room, extra time). Validity and comparability hold.

A **modification** changes what is measured (how the content is presented, number of items, format of response) to meet the unique needs of learners. Results may not be compared without additional equating exercises.

KNOW BEFORE YOU GO pretest, pilot and operational data collection

	Pretest	Pilot	Operational
# of test takers	12–50 learners like the target group	Larger convenience sample	Full sampled population
Goal	Find out if instructions, stimuli and timing work	Establish item quality incl. reliability, validity, logistics and equated forms	Capture the data that will be analyzed and reported
Changes	Revise freely	Revise with re-piloting as needed	No revisions; back up data

Equivalent forms (equal difficulty) are needed when the EGRA will be compared to other data points. Pilot multiple forms early.

03 IMPLEMENT THE EGRA field team, data, quality

- Roles:** assessors administer and run IRR pairs; supervisors handle school liaison, quotas and tablets; QCOs observe assessors and track issues. Assessors must be fluent in the assessment language.
- Training:** at least 5–6 days, more for two languages or extra surveys. Include a safeguarding session and dry runs.
- Pace:** 15–20 minutes per learner. A team of three does about 9–10 per hour.
- Materials:** for electronic assessment - one charged tablet per assessor, 1–2 spares per team, laminated stimuli, power banks; paper back-ups.
- Daily:** IRR on the first learner, upload data, supervisor report. Learners can stop at any time and it must be respected.

COMMON PITFALLS

Skipping adaptation of an old tool · too many subtasks or languages · IRB approval left to the end · one long passage instead of two · modifications that compromise comparability · revising the tool mid-collection without anchor items.

ASSESSMENT QUALITY THRESHOLDS

Parameter	What this means
Offer enough items for learners to demonstrate their ability	Offer items that measure precursor skills aligned to the Global Proficiency Framework, totaling to at least 20 score points with at least 10 of them on comprehension.
Ensure variety and sufficiency of items to measure Comprehension	Offer 10 or more items that measure direct and inferential comprehension appropriate for the grade level assessed. Develop rich and relevant passages of different topics and genres. Avoid one long passage.
Align the assessment to the curriculum and where children actually are	Work with experts—language, curriculum, reading experts in the country/context being assessed. Triangulate between items, GPF constructs, curriculum and classroom instruction to land on items that will accurately measure performance.
Test items to ensure they meet item quality parameters	During pilot retain only items that meet difficulty of 0.20–0.90; item-total correlation at least 0.20; discrimination suited to the population. Cronbach's alpha at least 0.80 on the comprehension and subtasks.
Select sample to ensure national representation (if applicable)	Work from a reliable and recent population frame and select a sample based on 95% CI within 5 percentage points. Document the frame, design effect, exclusions, weights, substitution.
Always pilot test the items	Pilot at least 2–3 complete versions of the test items so the retained items meet all parameters above. Enumerator training can also be piloted.
Balance a thorough test with what is age-appropriate	Use zero score imputation only where pilot-tested skip patterns justify it; score out of all items. To reduce learner fatigue and frustration in young test-takers, imputing zero scores on a subtask may be warranted, based on pilot data.
Recognize that data quality relies on training and QC	Operational data collection relies on high-quality training that uses reproducible training manual, has training completion requirements, uses IRR plans, tracks and reports response rates, incidents and resolutions daily, and manages data from classroom to server under a data security plan.

04 ANALYZE

Led by a psychometrician or statistician; the Manager asks guiding questions. Decide on CTT vs IRT/Rasch, whether to equate forms, apply weights, or drop cases and items.

- Level 1** Set up and clean data.
- Level 2** Score, build variables and weights.
- Level 3** Descriptive and statistical analysis and output tables.

05 REPORT

Plan documentation from Design onward: a test development report, sampling report, technical annex tables and benchmarking report. Fit the format to the audience and hold data dives before final reporting. Compare across languages or contexts by percent meeting a common benchmark.